Theor Appl Genet (1999) 98: 126–134

Y.-K. Lee · M. Ciaffi · R. Appels · M. K. Morell

The low-molecular-weight glutenin subunit proteins of primitive wheats. II. The genes from A-genome species

Received: 24 November 1997 / Accepted: 18 August 1998

Abstract Three accessions of *T. boeoticum* were selected for the cloning and sequencing of novel low-molecular-weight glutenin subunit (LMW-GS) genes, based on the results of SDS-PAGE and PCR analyses of the LMW-GS diversity in A-genome wheat (Lee et al. 1998 a). A comparison of the nucleotide and deduced amino-acid sequences of three cloned genes, LMWG-E2, LMWG-E4 and LMWG-AQ1, both to each other and to other known LMW-GS genes was carried out. The N-terminal domains showed one variable position; GA**G** (coding for a glutamic acid) for the E-type, and GA**T** (coding for an aspartic acid) for the Q-type. The comparisons of the LMW-GSs in the literature and this paper define three different types of N-terminal sequences; M**ETS**CIPGLERPW and M**DTS**CIPGLERPW from the durum and A-genome wheats, and M**ETR**CIPGLERPW from the hexaploid and D-genome wheats. The repetitive domains were AC-rich at the nucleotide level and coded for a large number of glutamine residues; this region showed 16 variable positions changing 12 amino-acid residues, three triple nucleotide deletions/additions, a large deletion of 18 nucleotides in LMWG-E4 and a deletion of 12 nucleotides in LMWG-E2. In the C-terminal domains 26 variable positions were found and 12 of these mutations changed amino-acid residues; no deletions/additions were present in this region. It was shown that the LMWG-E2 and LMWG-E4 genes could be expressed in bacteria and this allowed the respective protein products to be related back to the proteins defined as LMW-GSs in vivo.

Communicated by B. S. Gill

Y.-K. Lee · M. Ciaffi · R. Appels (✉) · M. K. Morell
CSIRO-Plant Industry, PO Box 1600, Canberra, ACT 2601, Australia

Key words Low-molecular-weight glutenin subunit proteins · Gene sequence · Expression in bacteria · A genome wheats

Introduction

The molecular characterisation of the wheat storage proteins provides a basis for determining the relationships between the molecular structure of the proteins and their functional roles in dough quality (reviewed in Shewry et al. 1994). A number of full-length high-molecular-weight glutenin subunit (HMW-GS) and gliadin genes have been isolated, and the molecular structures of these proteins have been deduced on the basis of their gene sequences (reviewed in Shewry et al. 1994). Only one subgroup of low-molecular-weight glutenin subunit (LMW-GS) gene sequences is currently available, the C-subunits (Okita et al. 1985; Colot et al. 1989; Cassidy and Dvorak 1991; D'Ovidio et al. 1992; Ciaffi et al. 1998), and none have been isolated from A-genome diploid wheats. The coding sequences of the LMW-GS genes are not interrupted by introns, so it is possible to clone these genes directly from genomic DNA and deduce their amino-acid sequences from the nucleotide sequences. The isolated genes can also be used for a further study of protein structure-function relationships through the expression of polypeptides in bacteria.

Three accessions of *Triticum boeoticum* were selected for cloning and sequencing of novel LMW-GS genes, based on the results of the SDS-PAGE and PCR analyses of Lee et al. (1998 a). These three accessions showed a relatively large number of unique LMW-GS bands in SDS-PAGE gels; the three accessions also showed diversity at the DNA level, as shown by the presence of several different LMW-GS PCR bands in appropriate experiments. Therefore, these three accessions were suitable candidates for obtaining novel

LMW-GS genes. This manuscript describes the cloning and sequencing of some of the novel LMW-GS genes and provides a comparison of the nucleotide and deduced amino-acid sequences of the cloned genes to each other and to other known LMW-GS genes. Two of the characterized genes could be expressed in bacteria and this allowed the protein products encoded by the genes to be related back to the proteins defined as LMW-GSs in vivo.

Materials and methods

Preparation of DNA

PCR was employed for isolating LMW-GS genes from genomic DNA of *T. boeoticum* accessions AUS 15823, AUS 90405 and AUS 90406. A pair of primers, DOV-1 (5′-AATTCATGAAGACCTTCC-TCGTCTT-3′) and DOV-3 (5′-AATTCGTAGGCACCAA CTCC-GGTGC-3′), were used to amplify entire LMW-GS genes including the region coding for the signal peptide. The primer sequences were based on the LMW-GS gene sequence from Colot et al. (1989). PCR conditions were the same as previously described by Lee et al. (1998 a). PCR products were electrophoresed in 1.2% low-melting-point agarose gels, excised under UV light, and purified using a QIAEX II DNA extraction kit from QIAGEN.

A ligation reaction was set up employing a pGEM-T System purchased from Promega and using 70 ng of purified PCR fragments and 50 ng of pGEM-T vector, following standard procedures. Transformation of *Escherichia coli* DH5a also followed standard procedures using a Bio-Rad Gene Pulser.

The procedure of colony hybridisation described by Grunstein and Hogness (1975) and Grunstein and Wallis (1979) was used to select plasmids containing LMW-GS gene sequences. The gene sequence in pTdUCD1 (Cassidy and Dvorak 1991) was labelled with ^{32}P (Megaprime DNA labelling system, Amersham) and used in the hybridization experiments.

Small-scale plasmid DNA preparations were carried out using standard procedures (Maniatis et al. 1989). The restriction endonucleases *Pvu*II, *Nde*I and *Bam*HI were employed for confirming the identity of the respective recombinant plasmid DNAs. For rapid screening of positive colonies, single colonies were re-streaked on a fresh plate and incubated overnight. Several colonies were transferred together into an Eppendorf tube, re-suspended in 13 µl of colony cracking buffer (50 mM NaOH, 5 mM EDTA, 0.5% SDS and 0.025% bromocresol green), vortexed thoroughly and incubated at 65°C for 40 min. The sample was mixed with 1.5 µl of 10 × DNA loading buffer (0.25% Bromophenol Blue, 0.25% Xylene Cylanol FF and 30% glycerol) and analysed in a 1% agarose gel.

Sequencing of the LMW-GS genes

Sequencing was carried out using an ABI Prism Dye Terminator sequencing kit purchased from Perkin Elmer, employing standard preparations of plasmids. The primers used for sequencing the 5′ and 3′ regions of the genes were either the Reverse primer (5′-CAG-GAAACAGCTATGACC-3′), or the Forward primer (5′-TGTA-AAACGACGGCCAGT-3′). On the basis of the sequences of the 5′ and 3′ regions, several primers were designed to sequence the internal part of the genes. The primers for internal sequencing are listed below:

LG-N 5′-AGATGCATCCCTGGTTTGGAG-3′,
LG-C 5′-AATGGAAGTCATCACCTCAAG-3′,
LG-repeat 5′-GAGGAATACCTTGCATGGGTT-3′,
LG-cdo 5′-AACCCATGCAAGGTATTCCTC-3′.

GCG seq!uence-analysis software (Genetics Computer Groups Incorporated, Wisconsin) was used for gene-fragment assembly and sequence alignment; alignments were checked manually and adjusted as required. The multiple sequence-alignment program, PileUp, was employed to plot a dendrogram showing the clustering relationships used to determine the order of the pairwise sequence alignments that produce the final multiple sequence alignment.

Bacterial expression of low-molecular-weight glutenin subunits

The bacterial expression vector pET-11a (Novagen) was used to express the genes in *E. coli*; the gene of interest was inserted using the *Nde*I and *Bam*HI sites. The LMW-GS genes were initially cloned into the pGEM-T vector after amplification using the PCR primers LG-SY, LG-SYD and LG-STOP. The DNA sequences encoding the signal peptides were excluded from the region amplified so as to obtain only the coding region for the mature LMW-GSs.

*Nde*I
LG-SY 5′-GGAATTCCATATGGAGACTAGCTGCATCCCTG-
GTTT-3′

*Nde*I
LG-SYD 5′-GGAATTCCATATGGATACTAGCTGCATCCCT-
GGTTT-3′,

*Bam*HI
LG-STOP 5′-CGCGGATCCTTAGTAGGCACCAACTCCGGT-
GGC-3′.

The primers LG-SY and LG-SYD were designed to anneal to the 5′ end of the E-type and the Q-type genes, respectively. These primers contained a novel *Nde*I site in the methionine codon at the mature N-terminus and included seven additional nucleotides at the 5′ end of the primers to allow for efficient restriction-enzyme digestion. The primer, LG-STOP, was designed to anneal to the 3′ end of both types of the gene, and included a novel *Bam*HI site and three additional nucleotides at the 5′ end of the primer. PCR reactions and subsequent digestion with *Nde*I and *Bam*HI employed standard procedures. Ligation reactions and transformation into an expression host (*E. coli* strain, JM109-DE3, Novagen) followed the suppliers instructions.

To express the proteins in bacteria, a fresh transformant containing the recombinant plasmid DNA and a negative control (the same strain bearing an empty pET-11a plasmid) were each inoculated into 5 ml of LB/ampicillin medium and incubated for 4 h. For the protein-expression assay, 1 ml of each culture was removed before induction, and the remaining culture was incubated for another 3 h in the presence of isopropyl β-D-thiogalactoside (IPTG) at a final concentration of 1.5 mM. The cells from the 1-ml cultures before and after induction were collected by centrifugation at 14 000 g for 5 min. The cell pellets were re-suspended in 60 µl of 55% isopropanol in the presence or absence of 10 µl of 1 M DTT, and incubated at 65°C for 30 min to extract the alcohol-soluble cell proteins. The cell suspensions were then centrifuged, and the supernatants were transferred into new tubes and mixed with 70 µl of SDS buffer [0.125 M Tris, 4% (w/v) SDS, 30% (v/v) glycerol and 0.005% (w/v) bromophenol blue, pH 6.8]. When the total cell proteins were analysed, the cells from 1-ml cultures were collected, mixed with 60 µl of water and 70 µl of SDS buffer in the presence or absence of DTT. Standard SDS-PAGE was used to analyze the proteins produced. The maximum expression of the LMW-GS genes in bacteria was checked by a time-course assay.

Results

Cloning of the LMW-GS genes from *T. boeoticum*

The LMW-GS PCR products generated using the primers DOV-1 and DOV-3 were approximately 900 bp in size and were confirmed to be LMW-GS genes by carrying out a "nested" PCR using the pair of primers LMW-GL and LMW-GR (Lee et al. 1998a). The latter primers targeted only variable parts of the LMW-GS genes within the region amplified by DOV-1 and DOV-3. The PCR results showed three bands produced from the DNA template of AUS 15823, two from AUS 90405 and three from AUS 90406 (data not shown). The size differences of the PCR products were small, as expected from the proteins present in the accessions analyzed.

A total of 43 clones were found to contain full-length inserts derived from the three accessions of *T. boeoticum*; AUS 15823 (15 clones), AUS 90405 (15 clones), and AUS 90406 (13 clones). The 5′ regions of all 43 clones were sequenced.

Sequences of the LMW-GS genes

Initial sequencing of the 5′ region of the genes gave nucleotide sequences of the signal peptide, the mature N-terminus and most of the repetitive region. According to the differences in deduced amino-acid sequences of these regions the clones fell into two groups (Fig. 1); although the analysis applied to six clones, LG-AE1, LG-AE2, LG-AE3, LG-AE4, LG-AQ1 and LG-AQ2, only three are shown. In the mature protein sequences, the first site dividing the two groups was at position 35, in the start of the repetitive domain; one group contained a glutamic acid (E35) and the other group a glutamine (Q35). These groups will be referred as 'E-type' and 'Q-type', respectively. The signal peptide consisted of 20 amino-acid residues and contained a substitution at the 17th amino-acid position which was diagnostic for the two groups, the E-type always contained a threonine at position 17 and the Q-type always contained an alanine in this position. Further categorisation of the clones could be made on the basis of differences in the repetitive region within each group; the E-type was divided into two subgroups and the Q-type was separated into four subgroups (see Fig. 1). The DNA sources and the number of clones of each subgroup are shown in Table 1. The differences between clones within the Q-type involved the deletion or addition of an entire repeat unit (defined at the deduced amino-acid sequence level). In the E-type, deletions or additions affecting glutamine residues in the repetitive domain were observed. The sequence types found from this study were unique within each accession analyzed, except for the E-4 gene which was present in two accessions, AUS 15823 and AUS 90405. Among six subgroups, three, E-1, E-3 and Q-2, contained stop codons in the repetitive domain; therefore, these were excluded from further sequencing studies. All the stop codons occurred by point mutation, namely a conversion of **C**AA to **T**AA.

Representative clones of the three subgroups that did not contain stop codons in their coding regions were fully sequenced (Fig. 2). These genes were named LMWG-E2 (= LG-AE2), LMWG-E4 (= LG-AE4) and LMWG-AQ1, and their lengths were 894 bp, 888 bp and 909 bp, respectively. Deduced amino-acid sequences of the three full genes showed the three typical domains of the LMW-GSs; an N-terminal domain (13 amino acids), a repetitive domain (approximately 70 amino acids) and a C-terminal domain (193 amino acids). The lengths of the deduced polypeptides of the clones including the signal peptide, LMWG-E2, LMWG-E4 and LMWG-AQ1, were 298, 296 and 303 amino acids, respectively (see Fig. 1).

A comparison of the three genes at the nucleotide sequence level (see Fig. 2), indicated there were two point mutations in the signal peptide region – one of which did not change the amino-acid residue. The N-terminal domains showed one point mutation; GA**G** (coding for a glutamic acid) for the E-type, and GA**T** (coding for an aspartic acid) for the Q-type. The repetitive domains were AC-rich at the nucleotide level and coded for a large number of glutamine residues. This region showed 16 point mutations changing 12 amino-acid residues, three triple-nucleotide deletions/additions, a large deletion of 18 nucleotides in LMWG-E4 and a deletion of 12 nucleotides in LMWG-E2. In the C-terminal domains 26 point mutations were present and 12 of these mutations changed amino-acid residues. No deletions/additions were found in this region.

The three LMW-GSs (LMWG-E2, LMWG-E4 and LMWG-AQ1), chosen for expression in bacteria and for futher functional studies (Lee 1998 b), showed 35 variable amino-acid positions in their respective protein sequences; 14 changed polarity and another 14 variants altered the properties of the side chains. The three proteins differ in their levels of hydrophobicity and in thier charge differences.

Table 1 The DNA sources and the number of clones of each subgroup

Item	AUS 15823	AUS 90405	AUS 90406
LG-AE1	14		
LG-AE2			9
LG-AE3		6	
LG-AE4	1	3	
LG-AQ1		6	
LG-AQ2			4
Total	15	15	13

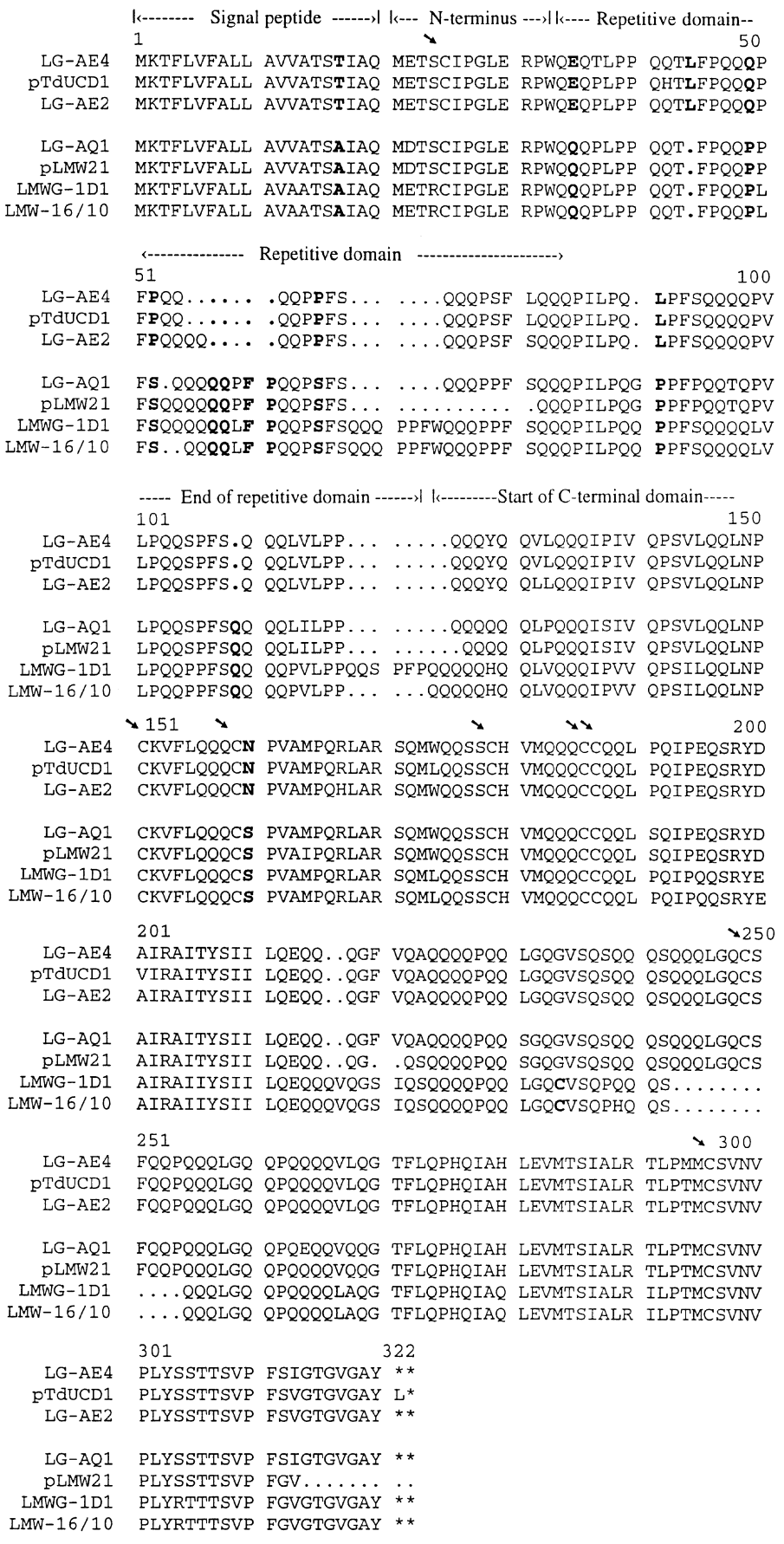
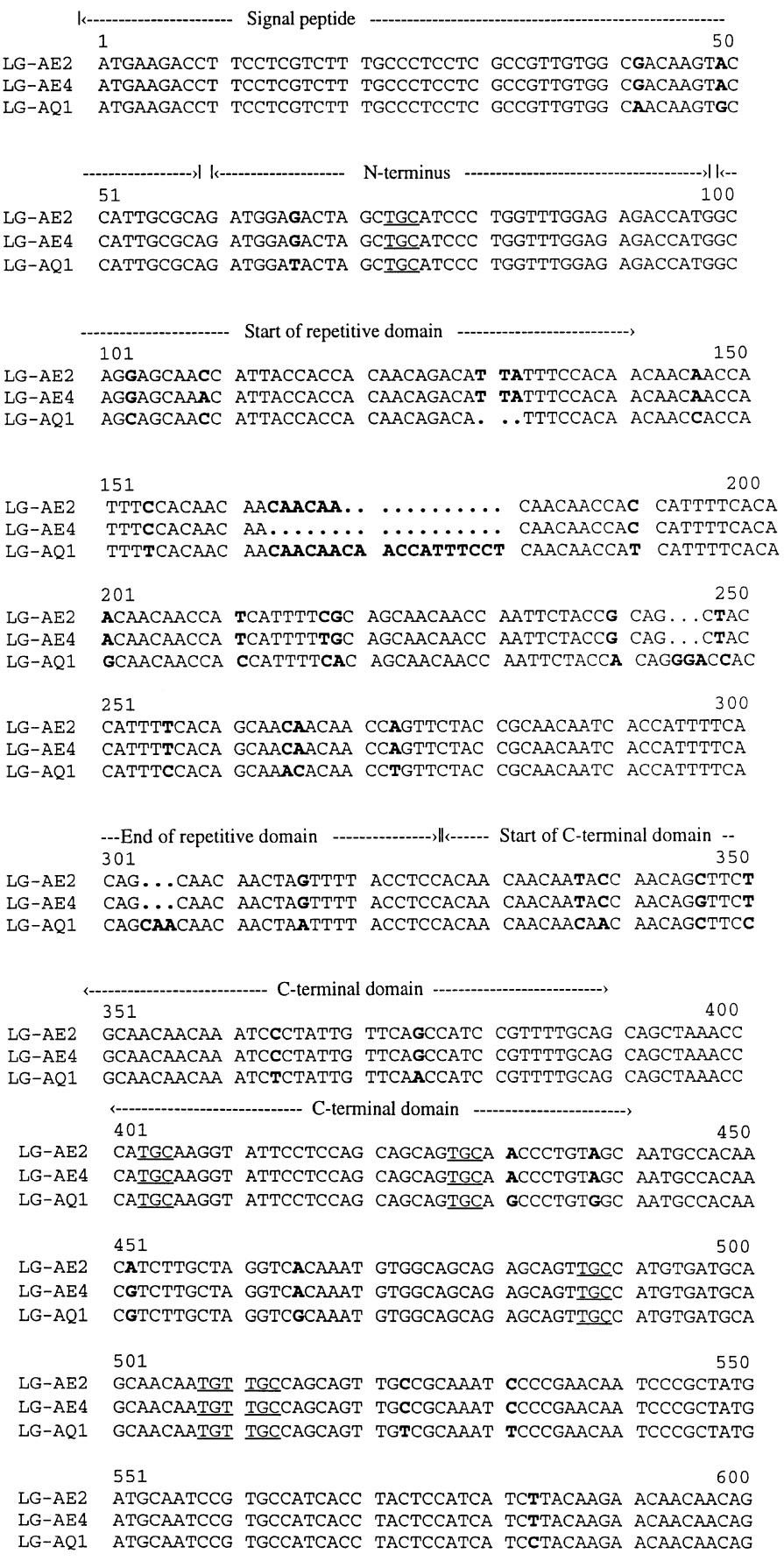


Fig. 1 Comparison of the deduced amino-acid sequences of the LMW-GS genes derived from A-genome diploid, D-genome diploid, tetraploid and hexaploid wheats. pTdUCD1 is a cDNA clone from tetraploid wheat, *T. durum*. pLMW21 is a PCR clone from tetraploid wheat, *T. durum*. LMWG-1D1 is a genomic DNA clone from the hexaploid wheat cultivar, Chinese Spring. LMW-16/10 is a genomic DNA clone of *T. tauschii*. LG-AE2, LG-AE4 and LG-AQ1 are PCR clones derived from the A-genome diploid wheat *T. boeoticum*. *Dots* were added as gaps for maximum homology between the sequences. The eight cysteine residues are marked with *arrows*. The residues at positions 17, 35, 44, 49, 52, 57, 58, 60, 61, 65, 91, 109 and 160 divide the E-type and Q-type sequences and are marked *bold*. An *asterisk* (*) indicates a stop codon. An *arrow* indicates the difference in the N-terminal region

Fig. 2 For legend see page 131



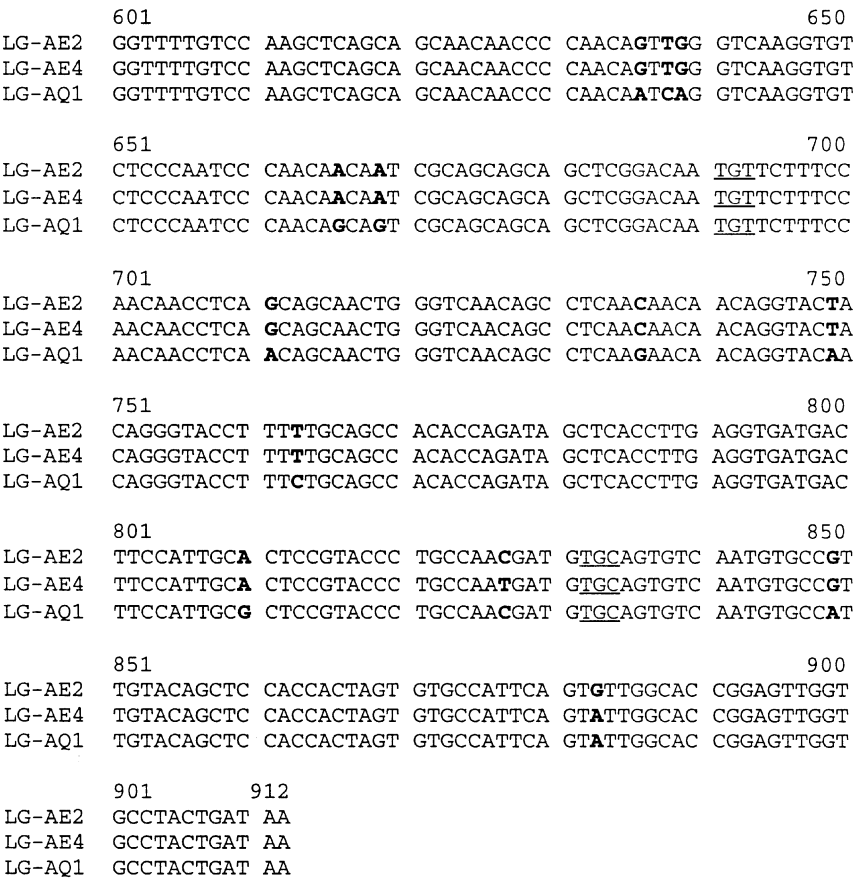


Fig. 2 A comparison of nucleotide sequences of the clones, LG-AE2, LG-AE4 and LG-AQ1. The nucleotides showing mutations are marked *bold*. The nucleotides (TGC and TGT) coding for cysteine residues are *underlined*. *Dots* are inserted as gaps for creating maximum homology between the sequences

A comparison of the genes at the deduced amino-acid sequence level indicated that the repetitive regions consisted of 11 repeat units, with 85% of the regions being dominated by four amino acids; namely, glutamine, proline, leucine and serine. The fourth repeat unit showed a major deletion/addition, and half of the repeat unit was missing in the E-type genes. The repeat units at the amino-acid sequence level within a given gene did not, in general, show high levels of similarity. Glutamine residues appeared as mono-, di- and tripeptides throughout the polypeptides, comprising 30% of the amino-acid content, and the glutamines were highly conserved (95%) in these polypeptides. There were eight cysteine residues, one in the N-terminus and seven in the C-terminus, and the positions of the cysteine residues are 100% conserved in these polypeptides.

The molecular weights of the polypeptides encoded by the three genes were calculated by the mean molecular weight of individual amino acids (D'Ovidio et al. 1994). The predicted molecular weights of the polypeptides encoded by LMWG-E2, LMWG-E4 and LMWG-AQ1 were 31.2 kDa, 30.9 kDa and 31.7 kDa, respectively.

Comparison of the amino-acid sequences of the LMW-GS genes

The deduced amino-acid sequences of LMWG-E2, LMWG-E4 and LMWG-AQ1 were compared to those of the four other LMW-GS gene sequences available; a cDNA clone, pTdUCD1, from the tetraploid wheat *Triticum durum* (Cassidy and Dvorak 1991); a PCR clone, pLMW21, from the tetraploid wheat *T. durum* (D'Ovidio et al. 1992); a genomic DNA clone, LMWG-1D1, from the hexaploid wheat cultivar Chinese Spring (Colot et al., 1989); and a genomic DNA clone, LMW-16/10, from the D-genome diploid wheat *Triticum tauschii* (Ciaffi et al. 1998). The seven amino-acid sequences compared shared 86% homology (see Fig. 1). Even including the amino-acid sequences from the different wheats, the separation between the Q-type and the E-type was obvious and was reinforced by parallel differences in amino-acid residues at positions 17, 35, 44, 49, 52, 57, 58, 60, 61, 65, 91, 109 and 160 in the full length amino-acid sequences (see Fig. 1). In particular, the sequences of LMWG-E4 and pTdUCD1 showed extremely high homology (98%). The Q-type genes were further separated into two groups. LMW-1D1

and LMW-16/10 shared high similarity. The sequences of LMWG-AQ1 and pLMW21 were also very similar.

All of the seven genes showed highly conserved hydrophobic signal peptides. Three different types of N-terminal sequences were found; M**ETS**CIPGLERPW and M**DTS**CIPGLERPW from the durum and A-genome wheats, and M**ETR**CIPGLERPW from the hexaploid and D-genome wheats. In the repetitive domains there were 13 different repeat units (see Fig. 1). Repeat unit 6 (QQQPPFW) was an addition in LMWD-1D1 and LMW-16/10. Repeat unit 13 (QQSPFP) was present only in LMWG-1D1. Repeat unit 7 (QQQPPFS) was deleted in pLMW21. The repeat unit QQPPFS was relatively common, but all of the repeat units were not conserved in respect of their sequence and length within a gene. The Q-type sequences showed variation by the clear deletion or addition of repeat units (see Fig. 1), while the E-type sequences showed smaller differences by glutamine residue additions/deletions. The C-terminal domains were highly conserved at the amino-acid sequence level and the positions of the eight cysteines were consistent, although the one cysteine residue located toward the C-terminus of LMWG-1D1 and LMW-16/10 was shifted by a 12 amino-acid deletion. This deletion was restored by the two repeat unit additions so that the length of the polypeptide was very similar to the other polypeptides.

The amino-acid sequence comparison showed that the genes derived from the A-genome wheat shared very high homology with the genes from the tetraploid wheats containing the A and B genomes, but much less with the genes derived from the D-genome of hexaploid wheat and the D-genome of diploid wheat.

Subcloning of the LMW-GS genes for bacterial expression

The full gene sequences coding for mature LMW-GSs were amplified by PCR, specifically excluding the signal peptide region. Polypeptides expressed from the LMWG-E2 and LMWG-E4 genes were readily identified on SDS-PAGE gels under reducing conditions by comparing the total cell proteins of induced bacteria (Fig. 3a, b). Maximum expression was observed from the culture of a 10-h incubation.; the LMWG-AQ1 gene could not be expressed successfully in bacteria presumably due to an unusual feature in the final structure of the expressed protein, and this was not investigated further.

The protein encoded by LMWG-E2 could be assigned to a single protein band within the native endosperm proteins (lanes 1–4, Fig. 3 a); their mobilities did not coincide precisely after alkylation, suggesting that the protein expressed in bacteria is not identical to the native protein in some feature of its structure. In the case of LMWG-E4, the mobility of the alkylated protein could be assigned to a band in the native protein analysis (lanes 5–8, Fig. 3 b) but there was a small difference in mobility in the non-alkylated protein analysis. The sensitivity of electrophoretic mobility to some feature of the structure of the proteins is emphasized by the much slower mobility of the proteins relative to the 31-kDa marker (the predicted size of the proteins). Both the LMWG-E2 and LMWG-E4 proteins had the

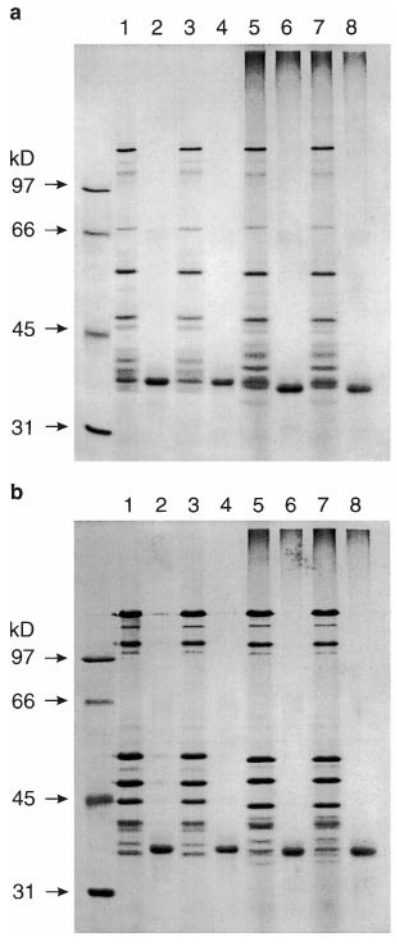


Fig. 3a, b Comparison of the bacterially expressed and endosperm synthesized LMW-GSs, using SDS-PAGE. The LMWG-E2 and LMWG-E4 genes were derived from *T. boeoticum* accessions AUS 90406 and AUS 90405 respectively. **a** Comparison of LMWG-E2 and the glutenin subunits from AUS 90406: *lane 1* AUS 90406 endosperm glutenin subunits extracted using 55% isopropanol/ DTT, *lane 2* LMWG-E2 extracted using 55% isopropanol/DTT, *lane 3* AUS 90406 endosperm glutenin subunits extracted using 10 M urea/DTT, *lane 4* LMWG-E2 extracted using 10 M urea/ DTT, *lanes 5–8* samples as in *lanes 1–4* but treated with an alkylating agent (4-vinylpyrimidine). **b** Comparison of LMWG-E4 and the glutenin subunits from AUS 90405: *lane 1* AUS 90405 endosperm glutenin subunits extracted using 55% isopropanol/ DTT, *lane 2* LMWG-E4 extracted using 55% isopropanol/DTT, *lane 3* AUS 90405 endosperm glutenin subunits extracted using 10 M urea/DTT, *lane 4* LMWG-E4 extracted using 10 M urea/DTT, *lanes 5–8* samples as in lanes 1–4 but treated with an alkylating agent (4-vinylpyrimidine)

N-terminal amino-acid sequences predicted from the structure of their respective genes.

Discussion

The comparisons of the amino-acid and nucleotide sequences of the seven representative LMW-GSs showed a clear structural organisation within the polypeptides. The N-termini were short and highly conserved. The repetitive domains contained several repeat units and showed major sequence variation. The C-terminal domains comprised over two-thirds of the polypeptide length with several stretches of glutamine residues and were relatively highly conserved. As LMW-GSs are synthesised on membrane-bound polysomes (Greene 1981) these proteins contained signal sequences that possessed a long stretch of hydrophobic residues. Different signal sequence lengths for the LMW-GSs have been determined by several workers (Kreis et al. 1985; Colot et al. 1989; Cassidy and Dvorak 1991). The lengths of the signal sequence of the genes isolated from the A-genome wheat in this study were assigned according to the results of direct N-terminal sequencing of the LMW-GSs (Tao and Kasarda 1989; Lew et al. 1992).

The deduced N-terminal amino-acid sequences of *T. boeoticum* show that they fit into the m-type subgroup of LMW-GSs as they all contain methionine as the first amino-acid residue (Tao and Kasarda 1989; Lew et al. 1992). The N-terminal sequence of LMWG-E2 and LMWG-E4, METSCIPGLERPW, was also found in a protein from a hexaploid cultivar, Yecora Rojo, by N-terminal sequencing (Lew et al. 1992), although the sequence (MDTSCIPGLERPW) of LG-AQ1 has not been found from N-terminal sequencing studies of any hexaploid wheat. As the durum wheat clone pLMW21 also contained a MDTSCIPGLERPW sequence in the N-terminal domain, this N-terminal sequence may be specific to the A-genome.

The repetitive domain shows clear variation in this group of proteins by the deletion and addition of residues and repeat units. This variation is a primary source of the polypeptide length polymorphisms. Its repetitive structure most likely facilitates rapid divergence by allowing slippage, leading to the duplication or deletion of sequences, during replication (Cassidy and Dvorak 1991). Although the individual repeat units are not conserved in their length and sequence within a gene, the irregular repeat units were consistent among the LMW-GSs and allows them to be distinguished from other wheat storage proteins.

Point mutations are commonly observed throughout the LMW-GS gene sequences, although half of them are silent or result in switching to other conserved amino acids; the point mutations are one of the key factors causing divergence within the LMW-GS group. Some of the point mutations result in amino-acid changes that can be predicted to cause changes in molecular size and charge. Furthermore, these changes would affect the secondary structure of the polypeptides and have the potential to modify the interaction between the polypeptides. This may be important in determining the roles of individual LMW-GSs in gluten structure.

The cysteine residues are highly conserved in the group of LMW-GSs analysed in this paper. This high level of conservation may reflect the importance of cysteine residues in stabilizing the secondary structure of the polypeptides (reviewed by Shewry and Tatham 1997). As a result, these cysteine-containing regions of the gene would evolve more slowly (Cassidy and Dvorak 1991). As the LMW-GSs give rise to polymers some of these cysteines are predicted to form intermolecular disulphide bonds, and thus determine the tertiary structure of the proteins (reviewed by Shewry and Tatham 1997). Subtle changes around the cysteine residues, such as the deletion near the 7th cysteine residue in the C-terminus of LMWG-1D1, and the substitutions near the 2nd cysteine of LG-AQ1 and the 8th cysteine residue in LMWG-E2, may influence the intermolecular disulphide bond arrangement of the proteins. As the number and positions of the cysteine residues are critical factors that affect dough properties (Shewry et al. 1995), it would be valuable to determine which of the cysteine residues are capable of forming the intermolecular disulphide linkages, and how the neighbouring amino acids are involved in making the linkage(s) feasible.

The studies in this paper identify the repetitive sequence domain of the LMW-GS proteins as a source of useful variation in gene structure that can be targeted for further work in assaying variation in hexaploid wheats (see also Van Campenhout et al. 1995). The studies also show that two of the three LMW-GS genes studied can be expressed in bacteria and used as a source of protein for functional analyses (Lee et al. 1998 b). Although the small differences in electrophoretic mobility between native and bacterially expressed proteins indicate that not all features of structure (possibly relating to post-translation modifications of the native proteins) in the respective proteins are identical, the available information on the solubility and antibody binding characteristics (Ciaffi et al. 1998) indicate that many features of their secondary structure must be very similar.

Acknowledgements This work was supported by a GDRC PhD fellowship to Y.K.-L. The input by Dr. R. Gupta into the project in its early stages is acknowledged.

References

Cassidy BG, Dvorak J (1991) Molecular characterisation of a low-molecular-weight glutenin cDNA clone from *Triticum durum*. Theor Appl Genet 81:653–660

Ciaffi M, Lee Y-K, Tamas L, Gupta R, Skerritt J, Appels R (1999) The low-molecular-weight glutenin subunit proteins of primitive wheats. III. The genes from D genome species. Theor Appl Genet 98 : 135–148

Colot V, Bartels D, Thompson R, Flavell R (1989) Molecular characterisation of an active wheat LMW glutenin gene and its relation to other wheat and barley prolamin genes. Mol Gen Genet 216 : 81–90

D'Ovidio R, Tanzarella OA, Porceddu E (1992) Nucleotide sequence of a low-molecular-weight glutenin from *Triticum durum*. Plant Mol Biol 18 : 781–784

D'Ovidio R, Porceddu E, Lafiandra D (1994) PCR analysis of genes encoding allelic variants of high-molecular-weight glutenin subunits at the *Glu-D1* locus. Theor Appl Genet 88 : 175–180

Greene FC (1981) In vitro synthesis of wheat (*Triticum aestivum* L.) storage proteins. Plant Physiol 68 : 778–783

Grunstein M, Hogness DS (1975) Colony hybridisation. Proc Nat Acad Sci USA 120 : 3961–3965

Grunstein M, Wallis J (1979) Colony hybridisation. In: Wu R (ed) Methods in enzymology. Academic Press, New York, pp 379–389

Kreis M, Shewry PR, Forde BG, Forde J, Miflin BJ (1985) Structure and evolution of seed storage proteins and their genes with particular reference to those of wheat, barley and rye. Oxford Surveys Plant Mol Cell Biol 2 : 253–317

Lee Y-K, Bekes F, Appels R, Morell M (1999a) The low-molecular-weight glutenin subunit proteins of primitive wheats. I. Variation in A-genome species. Theor Appl Genet 98 : 119–125

Lee Y-K, Bekes F, Gras P, Ciaffi M, Appels R, Morell M (1999b) The low-molecular-weight glutenin subunit proteins of primitive wheats. IV. Functional properties of products from individual genes. Theor Appl Genet 98 : 149–155

Lew EJ-L, Kuzmicky DD, Kasarda DD (1992) Characterisation of low-molecular-weight glutenin subunits by reversed-phase high-performance liquid chromatography, sodium dodecyl sulfate-polyacrylamide gel electrophoresis, and N-terminal amino acid sequencing. Cereal Chem 69 : 508–515

Maniatis T, Sambrook J, Fritsch EF (1989) Molecular cloning: a laboratory manual. Cold Spring Harbour Laboratory, Cold Spring Harbor, New York

Okita TW, Cheesbrough V, Reeves CD (1985) Evolution and heterogeneity of the a-/b-type and g-type gliadin DNA sequences. J Biol Chem 260 : 8203–8213

Shewry PR, Tatham AS (1997) Disulfide bonds in wheat gluten proteins. J Cereal Sci 25 : 207–227

Shewry PR, Miles MJ, Tatham AS (1994) The prolamin storage proteins of wheat and related cereals. Prog Biophys Mol Biol 61 : 37–59

Shewry PR, Tatham AS, Barro F, Barcelo P, Lazzeri P (1995) Biotechnology of breadmaking: unravelling and manipulating the multi-protein gluten complex. Bio/Technology 13 : 1185–1190

Tao HP, Kasarda DD (1989) Two-dimensional gel mapping and N-terminal sequencing of LMW-glutenin subunits. J Exp Bot 40 : 1015–1020

Van Campenhout S, Vander Stappen J, Sagi L, Volckaert G (1995) Locus-specific primers for LMW glutenin genes on each of the group-1 chromosomes of hexaploid wheat. Theor Appl Genet 91 : 313–319